The Development of a CSD Subset: A Collection of Metal-Organic Frameworks for Past, Present and Future

Peyman Z. Moghadam,[a,†,*] Aurelia Li,[a,†] Seth B. Wiggin,[b] Andi Tao,[a] Andrew G. P. Maloney,[b] Peter A. Wood,[b] Suzanna C. Ward,[b] and David Fairen-Jimenez[a,*]

[a]Adsorption & Advanced Materials Laboratory (AAML), Department of Chemical Engineering & Biotechnology, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, UK

[b] The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, UK

*Correspondence and requests for materials should be addressed to Peyman Z. Moghadam (pzm20@cam.ac.uk) or to David Fairen-Jimenez (df334@cam.ac.uk); website: http://people.ds.cam.ac.uk/df334

† These authors contributed equally

ABSTRACT: We report the generation and characterization of the most complete collection of metal-organic frameworks (MOFs) maintained and updated, for the first time, by the Cambridge Crystallographic Data Centre (CCDC). To set up this subset, we asked the question "what is a MOF?" and implemented a number of "look-for-MOF" criteria embedded within a bespoke Cambridge Structural Database (CSD) Python API workflow to identify and extract information of 69,666 MOF materials. The CSD MOF subset is updated regularly with subsequent MOF additions to the CSD, bringing a unique record for all researchers working in the area of porous materials around the world, whether to perform high-throughput computational screening for materials discovery or to have a global view over the existing structures in a single resource. Using this resource, we then developed and used an array of computational tools to remove residual solvent molecules from the framework pores of all the MOFs identified and went on to analyze geometrical and physical properties of non-disordered structures.

INTRODUCTION

Metal-organic frameworks (MOFs)[1-3] are one of the most exciting recent advances in porous materials science. MOFs are crystalline materials containing metal clusters connected by organic linkers (Figure 1 inset) and are characterized by their wide range of well-defined and, in principle, tailorable pore sizes (from micro- to mesoporosity), pore geometries, high void fractions and large surface areas. MOFs can reach apparent surface areas as high as 10,000 m²/g versus ca. 1000 m²/g for zeolites and up to 3500 m²/g for activated carbons,[4-5] and their large pore volumes have been unsurpassed by any other porous material to date. These unique properties have paved the way for MOF research to grow substantially and applications are being considered in many areas including; gas storage,[6-7], separation,[8-10], catalysis,[11-12] carbon capture[13-15] as well as biomolecule encapsulation[16], drug delivery[17-19] and imaging[20].

Because of the modular nature of MOFs, an almost unlimited number of structures can be envisioned. The Cambridge Crystallographic Data Centre (CCDC) collates and curates the Cambridge Structural Database (CSD) – the world's repository of small molecule crystal structures, which includes small organics as well as MOFs and other porous materials.[21] As shown in Figure 1, the number of entries in the CSD has substantially increased over the last 44 years, reaching a milestone of 850,000+ entries in 2016. Among these, the number of MOFs has also increased dramatically in the past decade to an estimated number of ca. 70,000 materials, vide infra. This trend is only going to increase even further as the building-block approach to MOF synthesis opens up the possibility to create more and more materials. This clearly creates multiple opportunities in their applicability, but it also creates the following challenge: *how does one identify the most promising structures, among the thousands of possibilities, for a particular application?* Although the demand for novel adsorbents is urgent, experimental discovery is bound by high costs and time-consuming procedures amongst structures dispersed throughout the literature.
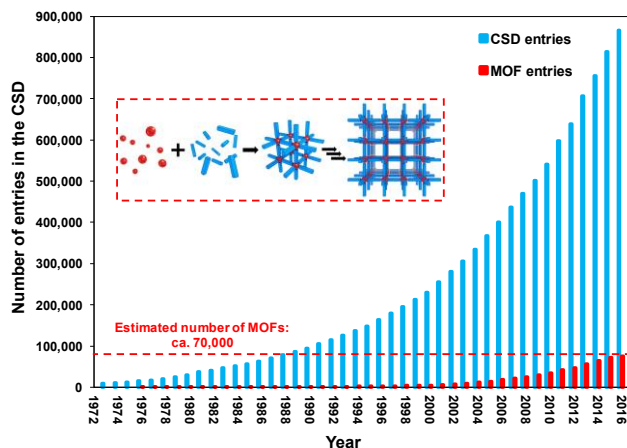
Figure 1. Growth of the CSD and MOF entries since 1972. The inset shows the MOF self-assembly process from building blocks: metals (red spheres) and organic ligands (blue struts).

Even though all published MOF structures are collected in the CSD, there is currently no easy way of distinguishing them from the rest of the structures in the CSD. Indeed, the question of "what is a MOF?" remains to this day debatable amongst MOF scientists

and, as recently described by IUPAC (International Union of Pure and Applied Chemistry)[22], at times varies depending on the application for which they are being studied. In a nutshell, finding and extracting MOF structures in the CSD manually is complex, and will only become more difficult as the number of published structures increases.

Several research groups have built MOF, zeolitic imidazolate framework (ZIF) or porous polymer network (PPN) databases using computer-based approaches to make hypothetical structures[6, 23-25]. For example, Wilmer et al.[23] built a hypothetical MOF database containing 137,953 structures generated from the recombination of a library of 102 building blocks containing secondary building units (SBUs) and organic linkers from available crystallographic data of existing MOFs. Though very useful in predicting structure-property relationships for a variety of adsorption applications[26-29] as well as identifying promising structures, the main drawback of a hypothetical database is the necessity to design a synthesis protocol after the top structures are identified, which is often complex, if not impossible. An alternative is given by databases built from experimental crystal data, such as those gathered in the CSD. Different research groups have previously created MOF databases based on experimental structures directly extracted from the CSD. In 2012, Watanabe et al.[30] extracted 30,000 extended metal-organic compounds from the CSD from which 1,167 3D MOF materials were analyzed for CO_2/N_2 separation. In 2013, Goldsmith et al.[31] generated 22,700 "computation-ready" 3D MOFs from a total of 38,800 metal-organic compounds from 550,000 CSD structures and studied them for hydrogen storage. The most recent MOF database was the first publicly available database developed by Chung et al.[32] in 2014. The database, named "Computation-Ready, Experimental MOFs" (CoRE MOF) database, contains over 4,700 3D MOFs, cleaned from solvents with a pore limiting diameter (PLD) larger than 2.4 Å. All three studies focused on gas adsorption, storage and separation applications and follow roughly the same protocol: extraction of a first set of structures from CSD; removal, when the process is not too complex, of bound and/or unbound solvents and guest molecules; followed sometimes by the elimination or repair of disordered structures and the characterization of structural properties.

The above CSD-derived databases have proved to be very useful. However, since they are not integrated within the CSD, they require manual updates whenever the CSD is updated and therefore, subsequent additions of MOF structures are not accounted for. Also, the structures in the aforementioned databases were targeted for adsorption applications, therefore only 3D structures with appropriate pore and window size were included depending on the kinetic diameter of the desired adsorbate. In addition, there are also issues with implementing too general or too restrictive search criteria for MOFs. For example in the Goldsmith et al. work, eliminating MOFs with alkali metal ions is too restrictive, as there are known examples of alkali-based MOFs.[33] In the CoRE MOF database, a small number of non-MOF structures (see Figure S1) are found, and MOF structures, such as FMOF-1 (Figure S2), are noted as missing. Although the missing structures can be explained by their elimination due to disorder reasons or small pore window diameters, the presence of non-MOF structures can only be caused by issues in the definition of potential MOFs during the first step of MOF identification. Apart from that, previous attempts to create MOF databases from the CSD have removed bound solvent molecules. Although this is expected for well-known metal clusters such as the Cu-Cu paddlewheel found in e.g. HKUST-1, or the CPO-27/MOF-74 family, in some cases, MOFs contain crystallographic solvents that cannot be successfully removed without provoking a degree of pore or crystalline structure collapse.[34]

To overcome these challenges, we aim to use a multitude of computational methods rooted in automated database construction and intelligent data mining. Here, we present the generation – for the first time – of a curated, regularly updated CSD MOF subset containing all the MOF materials published so far, and maintained by the CCDC. We expect that the CSD MOF subset will be of significant interest to a diverse group of researchers around the world enabling access and intelligent probing of MOF materials in a single updateable database for fast materials screening and discovery in a wide range of applications. The MOF subset described here has a number of advantages over previous studies. First, the database has been integrated into the CSD as a subset, allowing direct searches and automatic updates with subsequent addition of MOF structures. Second, the identified MOFs are not restricted to adsorption applications, therefore 1D, 2D and 3D MOF and MOF-like structures with all kinds of pore sizes as well as non-porous structures are included. Third, users are able to carry out bond-type or cluster-type oriented searches of MOFs that should significantly speed up the screening of MOFs for different applications. This is the most complete collection of MOFs available and the only collection integrated within the CSD database and software suite. The intention is to provide the widest possible subset to enable research in a wide variety of applications for which the users of this resource are able to refine criteria further to target their particular specifications.

What is a MOF?

Although prior to this project the CSD did not include a focused MOF subset, the CCDC informally adopted a guideline for a MOF description,[21] which can be easily captured in the structure search program ConQuest.[35] This simplistic guideline of a MOF is [a material with] *"a transition metal bonded to an oxygen or nitrogen atom via a polymeric bond"*, where "polymeric bonds" are defined as "bonds between repeating units". However, this guideline presents two drawbacks. First, the restriction to transition metals is limiting and excludes all other metals that are effectively synthesized in some MOF structures. These metals are alkalis[36], rare earths[37] and metalloids[38], among others. Second, although most of the time metals in MOFs are bonded to organic linkers via at least one nitrogen or oxygen atom, this is not always the case. The current IUPAC definition for MOFs as included in the provisional recommendation is the following: [a MOF is] "a coordination polymer (or alternatively coordination network) with an open framework containing potential voids"[22], where a coordination polymer is "a coordination compound continuously extending in 1, 2 or 3 dimensions through coordination bonds" and a coordination network is a subset of the former: "a coordination compound extending, through coordination bonds, in 1 dimension, but with cross-links between two or more individual chains, loops or spirolinks, or a coordination compound extending through coordination bonds in 2 or 3 dimensions". Thus, MOFs are coordination polymers, with the condition of having *potential* porosity. The task group for IUPAC's project on defining MOFs even considers that, for a structure to be considered as a MOF, "no physical measurements of porosity or other properties are demanded *per se*",[39] which makes MOF distinction even more difficult. This grey area surrounding the presence of voids is due to the ongoing debates about the essence of MOFs. For instance, voids are essential in gas adsorption, separation and sometimes catalysis applications, which are by far the most studied applications for MOFs. However, other scientists look into aspects where 3D structures or the presence/accessibility to porosity is not essential such as electrical conductivity[40] or luminescent[20] and magnetic[41] properties.

Although the IUPAC definition of MOFs remains very general and debated, all MOF scientists agree on their building blocks: the metal-containing units, or secondary building units

(SBU), and the organic linkers. Here, instead of using one global definition for MOFs, as implemented before[31], we decided to focus on describing a variety of metal clusters and their linkers and on describing further their link. This inevitably involves more criteria, corresponding to the description of the various numbers of ways of connecting the metals and the organic linkers, but also prevents non-MOFs from being included in the database.
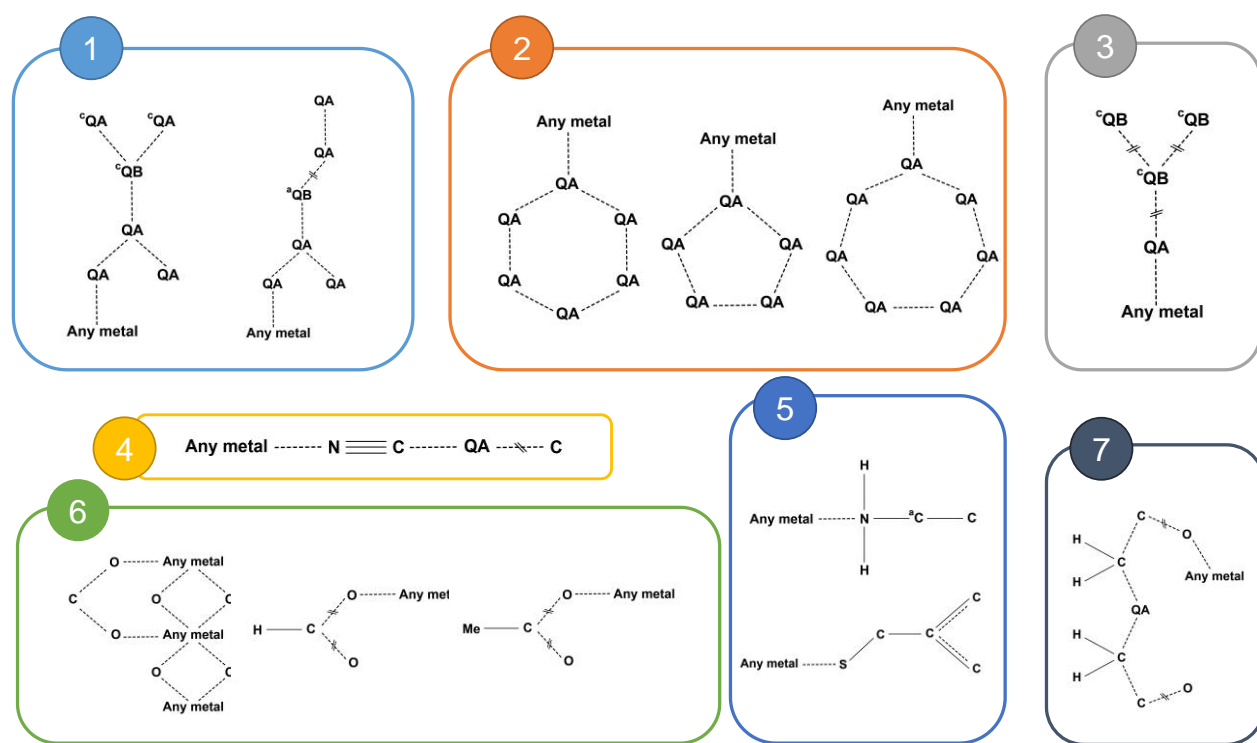
Building the CSD MOF subset

ConQuest[35] is the primary software for searching and retrieving information from the CSD, providing an extensive range of flexible search options. In our search query we focused on describing the link between the metal ion and the organic compound using different criteria. We systematically added the keywords "catena-", "catena(" and "catena[" in all searches; *catena* is a tag the CCDC uses for polymeric structures (i.e. any structure containing repeating units) containing metals; they are usually either followed by "-", "(" or "[". During the building process, the database was constantly compared to known existing MOFs and to the CoRE MOF database. Each time a known MOF structure not yet included in the CSD MOF subset was found in the literature, a new criterion for ConQuest was developed. The criteria were subsequently refined until the structures found by ConQuest included the set of known MOF structures.

Figure 2 shows the seven criteria we used for the building of the CSD MOF subset based on the version 5.37 of CSD including updates up to May 2016. These criteria are not mutually exclusive: some structures apply for several criteria. The union of the seven criteria yields 69,666 1D, 2D and 3D MOF candidates. Table S1 and Figure S3 represents the relative importance of the different criteria and their overlapping. It is worth mentioning that these criteria can be revised or added to if required in the future for the regular updates of the CSD MOF subset. The following section describes each criterion in detail.

We defined criterion **1** based on the metal-linker connection point that comprises the majority of MOFs synthesized today, i.e. carboxyl groups. Some of the well-known examples of this type include IRMOF-1, HKUST-1 and UiO-66 and their derivatives where in the simplest forms contain 1,4-benzenedicarboxylic (BDC) linkers connected to zinc-, copper- and zirconium oxide metal corners. Carboxylate-based linkers can be classified into three main groups summarized in Figure S4: those containing aromatic rings only, those containing hydrocarbon chains only, and those consisting of a combination of aromatic rings and hydrocarbon chains. Figure 3 explains the evolution of criterion **1** from a simple benzenedicarboxylate linker to a definition that would encompass all carboxylate-based MOFs with cyclic hydrocarbon linkers. In its simplest form, metals are connected to the BDC linkers (Figure 3a). In order to include carboxylate-derived linkers, the atoms of oxygen and carbon have been replaced by QA, which could be C, N, O, P, B, S or C (Figure 3b).

Figure 2. Summary of the seven criteria designed to build the CSD MOF subset, where QA = O, N, P, C, B, S. QB = N, P, B, S, C, superscripts "c" and "a" impose the corresponding atoms to be "cyclic" or "acyclic", respectively. Me denotes methyl groups. The dotted line refers to any of the bond types stored in the CSD (single, double, triple, quadruple, aromatic, polymeric, delocalized and pi). The dotted line with the two lines through indicates a variable bond type (i.e. two or more of the options above). In these cases, the variable type is single, double or delocalized. The union of these seven criteria combined with the presence of catena in the compound name found 69,666 MOF structures from version 5.37 of the CSD including updates up to May 2016.
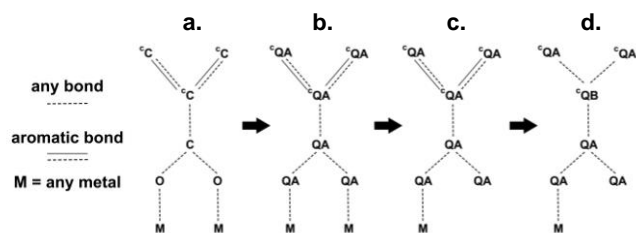
Figure 3. The evolution of criterion 1 to describe MOFs containing aromatic carboxylate-derived linkers. a. describes carboxylate linkers, b. and c. both describe carboxylate-derived linkers, c. yields results b. does not return, due to the way ConQuest carries out its searches. d. final criterion 1 extended to all types of bonds. QA = O, N, P, C, B, S. QB = N, P, B, S, C. Superscript "c" imposes the corresponding atom to be cyclic.

This search was further expanded to ensure the sub-structure search query was sufficiently generalized to include all MOF structures regardless of the choice of repeating unit in the CSD 2D representation. The sub-criterion shown in Figure 3c successfully captures these missing structures (see examples in Figure S6) in its 33,748 hits and also includes the structures described by Figure 3b. Next, the criterion is generalized to all types of rings and not limited to aromatic rings (Figure 3d). A similar procedure was carried out for MOFs with organic chains to cover both criterion 1a and 1b in Figure 1 (see Figure S9 for details). Criterion 1 provided the majority of the database, containing 50,046 structures, i.e. 74% of the final database.

Criterion **2** describes another significant number of structures containing cyclic organic compounds directly linked to at least one metal atom. Prominent examples of this type are some ZIFs and FMOF-1 type of structures, where *several* metal atoms are directly linked to the aromatic ring, although these rings do not necessarily have to be aromatic. Additionally, it is required for each ring to have at least one carbon atom in order to prevent inorganic cyclic structures such as the one shown in Figures S10. Criteria **2** hits 44,616 structures, i.e. 68% of the final database but 30,481 structures are overlapping with criterion **1**. More precisely, 13,616 new structures are brought into the database with criterion **2**. Together, criteria **1** and **2** yield 63,662 structures, i.e. 95% of the final database.

Criteria **3-7** build the remaining 5% structures of the CSD MOF subset and are focused on more specialized MOF materials with indirect metal-ring connections, linkers containing cyanides, metal atoms directly linked to nitrogen or sulfur atoms, formate-derived linkers and diglycolic acid-derived linkers, respectively. Full details for each criterion are given in the supporting information. Structures containing linkers composed of cyanides were not explicitly included in the database, as structures where carbon is only present as a cyanide do not currently meet the criteria for inclusion into the CSD. Combined together, criteria **1-7** yield a total of 69,666 structures in the CSD MOF subset. This subset was then analyzed to identify materials where the framework contained modelled disorder. The analysis showed that 21% of the *structures* contain disorder (including disordered or unmodelled solvent molecules), and 14 % of *frameworks* contain disorder (i.e. 9,549 MOF structures). To provide a collection of structures that can be used in high-throughput screening calculations, a second subset, the "CSD non-disordered MOF subset", was also created that excludes any structures where the framework was found to be disordered. This was not straightforward filtering process as use of the simple disorder filter would exclude structures containing any disorder at all, whether it is within the framework or the solvents.

These two lists are currently integrated into ConQuest as subsets of CSD entries. This not only ensures users access to all the newly published MOF materials regularly throughout the year but also provides a platform to perform targeted searches for different families of MOFs or to exclude structures with crystallographic disorder of the framework. These subsets will not only be updated as new MOF materials are published, but by using an API approach the subsets can evolve depending on changes to the definition of a MOF (i.e. the seven criteria described above) and subsets can be easily customized by users to match the criteria relevant to their particular area of interest.

Solvent removal from the CSD MOF subset

Before implementing computational techniques to characterize MOFs' textural properties, the residual solvent inside frameworks needs to be removed for adsorbate pore accessibility. Solvent is usually found in the frameworks due to the solvothermal synthesis. The elimination of solvent molecules from MOFs mimics the experimental "activation" process where the bound and unbound solvents are – desired to be – removed under heat and vacuum. To analyse solvents within the MOF structures, we used the library of 74 common solvent molecules included within Mercury. We found that 88% of the MOFs in the CSD MOF subset have solvent molecules. Among these, 52% have unbound and 48% have bound solvents. We found water to be by far the most common bound or unbound solvent, present in 47,478 MOF structures. Among the bound non-aqueous solvents, 33 of the 73 solvents from the CCDC list are present (see supporting information), with DMF being by far the most common one, present in 1509 MOF structures (Figure S23).

To remove the unbound solvent, the structures in CSD MOF subset were subjected to a cleaning process via a CSD Python API script (see supporting information) whereby only the heaviest component of each structure was retained, provided the heaviest component corresponds to the polymeric part. Removal of coordinated solvents to unsaturated metal centers is more challenging and has critical limits. For example, whilst the integrity of structures such as HKUST-1 or MOF-74 remains intact after the careful removal of metal-bound solvent molecules in experimental samples, the stability of some MOFs is critically dependent on the presence of metal-bound solvent molecules[28, 34]. These structures normally experience at least partial loss of porosity during the activation process. Moreover, since no further geometric optimization of frameworks are considered after the solvents are removed, the assumption to automatically remove them is not always correct. Our aim was to build a database that is representative of synthesized materials. Therefore, to overcome these issues and to take advantage of coordinately unsaturated metal sites in some of the well-known structures, we only removed the coordinated solvents for two types of MOFs: those containing copper-copper paddlewheels (e.g. HKUST-1) and MOF-74-type of metal clusters. These subsets were identified and then cleaned (i.e. solvent-removed) using a Python script in the following manner. After identifying the metal atoms in the heaviest part of the structure, all its bonds are removed. The multiple resulting components are then scanned through: if they are lone oxygen atoms (i.e. corresponding to partial representations of water or hydroxyls), or if they correspond to a known solvent from the solvent list provided by the CCDC they are removed from the structure, unless the solvent is bonded to multiple metal atoms. The rest of the structure is then rebuilt from the remnant components to form a *cleaned* structure. The in-house Python scripts developed in this project are available in the supporting information and can be

used for the removal of bound and unbound solvents for different classes of MOF materials. Users of these algorithms are advised to remove guest molecules in inorganic clusters where the elimination of coordinated solvents does not significantly affect the framework integrity.

Geometric and physical properties of the CSD MOF subset materials

The CSD MOF subset contains 69,666 structures with a wide range of chemical and structural properties. Having a priori knowledge of these properties is clearly useful for researchers to be able to scan through the existing materials and target specific structures with already-calculated textural properties such as surface area, pore volume, pore limiting diameter (PLD), the largest cavity diameter (LCD), void fraction, density as well as surface chemistry information. After the solvent removal procedure described above, each MOF structure present in the CSD MOF subset was characterized with the Zeo++ software package,[42] which uses Voronoi decomposition to identify probe-accessible regions of void space and calculate the accessible surface area, accessible volume, LCD, and PLD. Surface area calculations are performed using a probe of radius 1.86 Å, corresponding to N_2. Pore volume calculations used a probe with the radius set to zero. Covalent radii from the CSD were used for all MOF atoms.

Out of the 54,808 non-disordered MOF structures, 46,420 (ca. 85 % of the non-disordered CSD MOF subset) structures have gravimetric surface areas equal to zero, meaning that a N_2 probe sized molecule cannot access their pores for geometric surface area calculations. Figure 4 shows the distribution of the geometrical and physical properties of the 8,388 remaining structures. It is noteworthy to mention that all the structures have PLD values larger than 3.7 Å, ca. 3,600 more materials than previous MOF databases[32]. It can be seen that despite having a wide range of textural properties, the distribution of these properties are skewed towards certain limits with LCD and PLD values of < 15 Å and surface area values of < 2000 m^2/g. 55% of the 8,388 remnant structures have a gravimetric surface area between 500 and 2000

m^2/g and 61% have a volumetric surface area between 500 and 1700 m^2/cm^3. Similar to the CoRE MOFs, the majority of the structures have a density around 1 g/cm^3 and void fractions around 0.6. The geometric and physical properties for all the non-disordered CSD MOF subset structures can be found in the Excel file in the Supporting Information.

Conclusions and New Directions for the CSD MOF Subsets

We have developed seven "look-for-MOF" criteria to identify and extract MOF materials from a pool of >850,000 crystal structures in the Cambridge Structural Database (CSD). The constructed CSD MOF subset contains 69,666 already-synthesized MOFs and provides researchers access to a single, regularly updated database. This subset is currently integrated into the CCDC's structure search program ConQuest which allows for tailored structural queries (e.g. generation of MOF subsets based on secondary building units or selection of non-disordered materials) and visualization. We also developed and used an array of computational algorithms to first remove the solvent molecules from the CSD MOF subset and then to calculate the geometric and physical properties for all the non-disordered structures in the database.

To date, the CSD MOF subset is the most complete collection of MOF materials and is a major advance toward creation of a unique databank designated for past, present and future MOFs. Although the algorithms included here should allow the automatic identification of MOF structures, we propose to allow future MOFs to be flagged and confirmed by authors in the CCDC deposition service. From the characterization perspective, we will next focus on developing automated algorithms to address more challenging structures. This includes identification and manipulation of materials with crystallographic disorder or charge-balancing ions. The flexibility given by API scripts allow us to identify different classes of materials included in the CSD. Our future work will focus on the development of similar subsets for MOFs organic counterparts such as covalent organic frameworks (COFs), porous aromatic frameworks (PAFs) and porous organic polymers (POPs) among other subsets of MOFs.
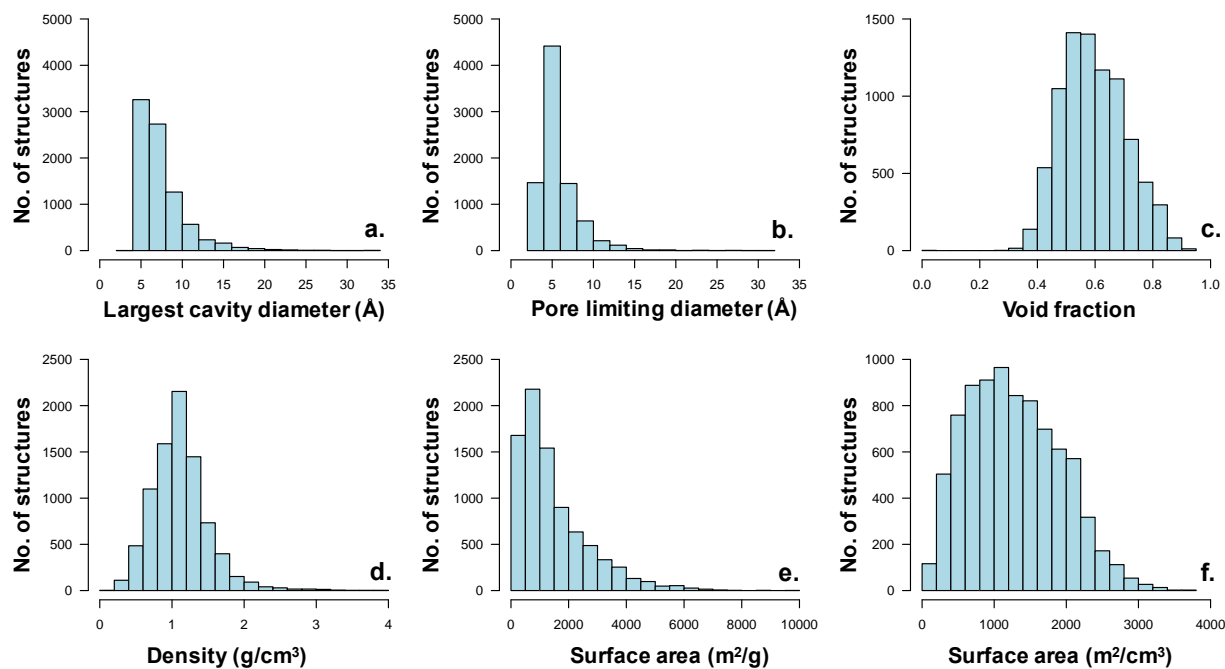
Figure 4. Histograms illustrating the geometric properties of 8,388 non-disordered MOFs with non-zero gravimetric surface area. a. largest cavity diameter, b. pore limiting diameter, c. void fraction, d. density, e. gravimetric surface area and f. volumetric surface area.

ASSOCIATED CONTENT

Supporting Information

Details of CSD MOF subset construction protocols, linked CSD refcodes and geometrical and physical properties for the structures in the subset, CCDC solvent list, Python scripts used for removing bound and unbound solvents. The cif files for all the CSD MOF subset structures are freely available from the CCDC's Access Structures service (https://www.ccdc.cam.ac.uk/structures); the subsets will be updated every quarter.

The Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Authors

Peyman Z. Moghadam (pzm20@cam.ac.uk), David Fairen-Jimenez (dff334@cam.ac.uk)

Author contributions

P.Z.M., S.B.W., A.G.P.M, P.A.W., S.C.W, N.F. and D.F-J. designed research; P.Z.M., A.L., A.T. and S.B.W. performed computational research; P.Z.M., A.L., S.B.W., P.A.W., S.C.W and D.F-J analyzed data; and P.Z.M., A.L., and D.F-J. wrote the paper.

Competing financial interests: D.F-J. has financial interest in the start-up company Immaterial Labs, which is seeking to commercialize metal-organic frameworks.

Biographies

Peyman Z. Moghadam is a Research Associate at the University of Cambridge. Prior to joining Cambridge, he was a postdoctoral fellow at Northwestern University from 2013-2015. He obtained his PhD in Chemical Engineering from the University of Edinburgh. His research is focused on the combination of data science, high-throughput molecular simulations, quantum chemical calculations and adsorption measurements to develop advanced functional materials in particular, self-assembled porous crystals for adsorption applications.

Aurelia Li obtained her MPhil in Advanced Chemical Engineering at the University of Cambridge and her Engineer's Degree at Ecole Centrale Paris.

Seth Wiggin obtained his PhD on framework materials from the University of Southampton in 2006. He is currently a Senior Scientific Editor at the CCDC.

Graduate student Andi Tao obtained his MEng in Engineering Science from University of Oxford.

Andrew Maloney is a Research and Applications Scientist at the CCDC. Before joining the CCDC he obtained his PhD in crystallography from the University of Edinburgh.

Peter Wood received his PhD in small molecule X-ray crystallography from the University of Edinburgh before joining the CCDC as a research scientist in 2007. Pete is now a senior research scientist and product manager of the CSD-System at the CCDC. He is focused on studying intermolecular interactions and understanding the organic and metal-organic crystalline solid state, both in research as well as through CSD software development.

Suzanna Ward is currently the manager of the Cambridge Structural Database at the CCDC. Before joining the CCDC in 2006 she obtained an MChem from the University of Southampton.

David Fairen-Jimenez received his PhD from the University of Granada, Spain, and completed postdoctoral studies at The University of Edinburgh and Northwestern University before joining Cambridge University as a Royal Society Research Fellow in 2012. In 2015 he was awarded a European Research Council (ERC) Consolidator Grant for the development of MOFs for drug delivery and bioimaging in cancer diagnosis and therapy.

REFERENCES

1.   Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M., Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **1999,** *402*, 276-279.

2.   Ferey, G., Hybrid porous solids: past, present, future. *Chem. Soc. Rev.* **2008,** *37*, 191-214.

3.   Horike, S.; Shimomura, S.; Kitagawa, S., Soft porous crystals. *Nat. Chem.* **2009,** *1*, 695-704.

4.   Farha, O. K.; Eryazici, I.; Jeong, N. C.; Hauser, B. G.; Wilmer, C. E.; Sarjeant, A. A.; Snurr, R. Q.; Nguyen, S. T.; Yazaydın, A. Ö.; Hupp, J. T., Metal–Organic Framework Materials with Ultrahigh Surface Areas: Is the Sky the Limit? *J. Am. Chem. Soc.* **2012,** *134*, 15016-15021.

5.   Gómez-Gualdrón, D. A.; Moghadam, P. Z.; Hupp, J. T.; Farha, O. K.; Snurr, R. Q., Application of Consistency Criteria To Calculate BET Areas of Micro- And Mesoporous Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2016,** *138*, 215-224.

6.   Gomez-Gualdron, D. A.; Colon, Y. J.; Zhang, X.; Wang, T. C.; Chen, Y.-S.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Zhang, J.; Snurr, R. Q., Evaluating topologically diverse metal-organic frameworks for cryo-adsorbed hydrogen storage. *Energy Environ. Sci.* **2016,** *9*, 3279-3289.

7.   Mason, J. A.; Oktawiec, J.; Taylor, M. K.; Hudson, M. R.; Rodriguez, J.; Bachman, J. E.; Gonzalez, M. I.; Cervellino, A.; Guagliardi, A.; Brown, C. M.; Llewellyn, P. L.; Masciocchi, N.; Long, J. R., Methane storage in flexible metal–organic frameworks with intrinsic thermal management. *Nature* **2015,** *527*, 357-361.

8.   Holcroft, J. M.; Hartlieb, K. J.; Moghadam, P. Z.; Bell, J. G.; Barin, G.; Ferris, D. P.; Bloch, E. D.; Algaradah, M. M.; Nassar, M. S.; Botros, Y. Y.; Thomas, K. M.; Long, J. R.; Snurr, R. Q.; Stoddart, J. F., Carbohydrate-Mediated Purification of Petrochemicals. *J. Am. Chem. Soc.* **2015,** *137*, 5706-5719.

9.   Li, J.-R.; Kuppler, R. J.; Zhou, H.-C., Selective gas adsorption and separation in metal-organic frameworks. *Chem. Soc. Rev.* **2009,** *38*, 1477-1504.

10.   DeCoste, J. B.; Peterson, G. W., Metal–Organic Frameworks for Air Purification of Toxic Chemicals. *Chem. Rev.* **2014,** *114*, 5695-5727.

11.   Liu, J.; Chen, L.; Cui, H.; Zhang, J.; Zhang, L.; Su, C.-Y., Applications of metal-organic frameworks in heterogeneous supramolecular catalysis. *Chem. Soc. Rev.* **2014,** *43*, 6011-6061.

12.   Corma, A.; García, H.; Llabrés i Xamena, F. X., Engineering Metal Organic Frameworks for Heterogeneous Catalysis. *Chem. Rev.* **2010,** *110*, 4606-4655.

13.   McDonald, T. M.; Mason, J. A.; Kong, X.; Bloch, E. D.; Gygi, D.; Dani, A.; Crocella, V.; Giordanino, F.; Odoh, S. O.; Drisdell, W. S.; Vlaisavljevich, B.; Dzubak, A. L.; Poloni, R.; Schnell, S. K.; Planas, N.; Lee, K.; Pascal, T.; Wan, L. F.; Prendergast, D.; Neaton, J. B.; Smit, B.; Kortright, J. B.; Gagliardi, L.; Bordiga, S.; Reimer, J. A.; Long, J. R., Cooperative insertion of

CO$_2$ in diamine-appended metal-organic frameworks. *Nature* **2015,** *519*, 303-308.

14. Sumida, K.; Rogow, D. L.; Mason, J. A.; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R., Carbon Dioxide Capture in Metal–Organic Frameworks. *Chem. Rev.* **2012,** *112*, 724-781.

15. Yazaydın, A. Ö.; Snurr, R. Q.; Park, T.-H.; Koh, K.; Liu, J.; LeVan, M. D.; Benin, A. I.; Jakubczak, P.; Lanuza, M.; Galloway, D. B.; Low, J. J.; Willis, R. R., Screening of Metal−Organic Frameworks for Carbon Dioxide Capture from Flue Gas Using a Combined Experimental and Modeling Approach. *J. Am. Chem. Soc.* **2009,** *131*, 18198-18199.

16. Li, P.; Modica, Justin A.; Howarth, Ashlee J.; Vargas L, E.; Moghadam, Peyman Z.; Snurr, Randall Q.; Mrksich, M.; Hupp, Joseph T.; Farha, Omar K., Toward Design Rules for Enzyme Immobilization in Hierarchical Mesoporous Metal-Organic Frameworks. *Chem* **2016,** *1*, 154-169.

17. Faust, T., Nanomedicine: MOFs deliver. *Nat. Chem.* **2015,** *7*, 270-271.

18. Bernini, M. C.; Fairen-Jimenez, D.; Pasinetti, M.; Ramirez-Pastor, A. J.; Snurr, R. Q., Screening of bio-compatible metal-organic frameworks as potential drug carriers using Monte Carlo simulations. *J. Mater. Chem. B* **2014,** *2*, 766-774.

19. Orellana-Tavra, C.; Baxter, E. F.; Tian, T.; Bennett, T. D.; Slater, N. K. H.; Cheetham, A. K.; Fairen-Jimenez, D., Amorphous metal-organic frameworks for drug delivery. *Chem. Commun.* **2015,** *51*, 13878-13881.

20. Miller, S. E.; Teplensky, M. H.; Moghadam, P. Z.; Fairen-Jimenez, D., Metal-organic frameworks as biosensors for luminescence-based detection and imaging. *Interface Focus* **2016,** *6: 20160027*.

21. http://www.ccdc.cam.ac.uk/.

22. Batten, S. R.; Champness, N. R.; Chen, X.-M.; Garcia-Martinez, J.; Kitagawa, S.; Ohrstrom, L.; O'Keeffe, M.; Suh, M. P.; Reedijk, J., Coordination polymers, metal-organic frameworks and the need for terminology guidelines. *CrystEngComm* **2012,** *14*, 3001-3004.

23. Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q., Large-scale screening of hypothetical metal–organic frameworks. *Nat. Chem.* **2012,** *4*, 83-89.

24. Lin, L.-C.; Berger, A. H.; Martin, R. L.; Kim, J.; Swisher, J. A.; Jariwala, K.; Rycroft, C. H.; Bhown, A. S.; Deem, M. W.; Haranczyk, M.; Smit, B., In silico screening of carbon-capture materials. *Nat. Mater.* **2012,** *11*, 633-641.

25. Martin, R. L.; Simon, C. M.; Smit, B.; Haranczyk, M., In silico Design of Porous Polymer Networks: High-Throughput Screening for Methane Storage Materials. *J. Am. Chem. Soc.* **2014,** *136*, 5006-5022.

26. Moghadam, P. Z.; Fairen-Jimenez, D.; Snurr, R. Q., Efficient identification of hydrophobic MOFs: application in the capture of toxic industrial chemicals. *J. Mater. Chem. A* **2016,** *4*, 529-536.

27. Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q., Structure-property relationships of porous materials for carbon dioxide separation and capture. *Energy Environ. Sci.* **2012,** *5*, 9849-9856.

28. Calahorro, A. J.; Salinas-Castillo, A.; Fairen-Jimenez, D.; Seco, J. M.; Mendicute-Fierro, C.; Gómez-Ruiz, S.; López-Viseras, M. E.; Rodríguez-Diéguez, A., Long lifetime photoluminescence emission of 3D cadmium metal–organic frameworks based on the 5-(4-pyridyl)tetrazole ligand. *Inorg. Chim. Acta* **2015,** *427*, 131-137.

29. Colón, Y. J.; Fairen-Jimenez, D.; Wilmer, C. E.; Snurr, R. Q., High-Throughput Screening of Porous Crystalline Materials for Hydrogen Storage Capacity near Room Temperature. *J. Phys. Chem. C* **2014,** *118*, 5383-5389.

30. Watanabe, T.; Sholl, D. S., Accelerating Applications of Metal–Organic Frameworks for Gas Adsorption and Separation by Computational Screening of Materials. *Langmuir* **2012,** *28*, 14114-14128.

31. Goldsmith, J.; Wong-Foy, A. G.; Cafarella, M. J.; Siegel, D. J., Theoretical Limits of Hydrogen Storage in Metal–Organic Frameworks: Opportunities and Trade-Offs. *Chem. Mater.* **2013,** *25*, 3373-3382.

32. Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q., Computation-Ready, Experimental Metal–Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014,** *26*, 6185-6192.

33. R. S. Forgan, R. A. S., J. J. Gassensmith, H. Furukawa, D. B. Cordes, Q. Li, C. E. Wilmer, Y. Y. Botros, R. Q. Snurr, A. M. Z. Slawin, J. F. Stoddart, Nanoporous Carbohydrate Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2012,** *134*, 406-417.

34. Howarth, A. J.; Peters, A. W.; Vermeulen, N. A.; Wang, T. C.; Hupp, J. T.; Farha, O. K., Best Practices for the Synthesis, Activation, and Characterization of Metal–Organic Frameworks. *Chem. Mater.* **2017,** *29*, 26-39.

35. http://www.ccdc.cam.ac.uk/solutions/csd-system/components/conquest/.

36. Forgan, R. S.; Smaldone, R. A.; Gassensmith, J. J.; Furukawa, H.; Cordes, D. B.; Li, Q.; Wilmer, C. E.; Botros, Y. Y.; Snurr, R. Q.; Slawin, A. M. Z.; Stoddart, J. F., Nanoporous Carbohydrate Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2012,** *134*, 406-417.

37. White, K. A.; Chengelis, D. A.; Zeller, M.; Geib, S. J.; Szakos, J.; Petoud, S.; Rosi, N. L., Near-infrared emitting ytterbium metal-organic frameworks with tunable excitation properties. *Chem. Commun* **2009,** 4506-4508.

38. Xie, M.-H.; Yang, X.-L.; Zou, C.; Wu, C.-D., A SnIV–Porphyrin-Based Metal–Organic Framework for the Selective Photo-Oxygenation of Phenol and Sulfides. *Inorg. Chem.* **2011,** *50*, 5318-5320.

39. Batten, S. R.; Champness, N. R.; Chen, X.-M.; Garcia-Martinez, J.; Kitagawa, S.; Öhrström, L.; O'Keeffe, M.; Suh, M. P.; Reedijk, J., Terminology of metal–organic frameworks and coordination polymers (IUPAC Recommendations 2013). *Pure Appl. Chem.* **2013,** *85*, 1715–1724.

40. Campbell, M. G.; Sheberla, D.; Liu, S. F.; Swager, T. M.; Dincă, M., Cu3(hexaiminotriphenylene)2: An Electrically Conductive 2D Metal–Organic Framework for Chemiresistive Sensing. *Angew. Chem. Int. Ed.* **2015,** *54*, 4349-4352.

41. Canepa, P.; Chabal, Y. J.; Thonhauser, T., When metal organic frameworks turn into linear magnets. *Phys. Rev. B* **2013,** *87*, 094407.

42. Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M., Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **2012,** *149*, 134-141.

43. Alezi, D.; Belmabkhout, Y.; Suyetin, M.; Bhatt, P. M.; Weseliński, Ł. J.; Solovyeva, V.; Adil, K.; Spanopoulos, I.; Trikalitis, P. N.; Emwas, A.-H.; Eddaoudi, M., MOF Crystal Chemistry Paving the Way to Gas Storage Needs: Aluminum-Based soc-MOF for CH$_4$, O$_2$, and CO$_2$ Storage. *J. Am. Chem. Soc.* **2015,** *137*, 13308-13318.

TOC FIGURE